DOCUMENT RESUME

ED 360 388 TM 020 396

AUTHOR Buechler, Mark
TITLE Performance Assessment. Policy Bulletin, No. PB-B13.
INSTITUTION Indiana Univ., Bloomington. Education Policy Center.
PUB DATE Apr 92
NOTE 7p.
AVAILABLE FROM Indiana Education Policy Center, Bloomington Office, Smith Center for Research in Education, Suite 170, Indiana University, Bloomington, IN 47405.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Accountability; Achievement Tests; Cost Effectiveness; *Educational Assessment; Educational Improvement; Elementary Secondary Education; *Multiple Choice Tests; *Portfolios (Background Materials); School Restructuring; Scoring; Standardized Tests; Student Evaluation; Test Construction; Testing Problems; *Test Use
IDENTIFIERS *Authentic Assessment; High Stakes Tests; *Performance Based Evaluation

ABSTRACT

At the root of the performance assessment movement is fairly widespread dissatisfaction with high-stakes multiple-choice tests. Many critics of multiple-choice tests argue that to improve instruction, tests themselves will have to improve. Hundreds of schools around the country are already experimenting with performance assessments, and many states are also experimenting with performance assessment. The states farthest along the road to performance assessment are Connecticut and Vermont. Vermont is the first state to use portfolios as part of a statewide assessment program. Among the advantages of performance assessments are the authenticity of what they say about what is assessed, the ways in which they offer students genuine intellectual challenges, and the opportunity they offer for restructuring schools. Problems with performance assessments are recognized in the areas of scoring, validity, instruction versus accountability, time constraints and teacher resistance, and cost. Those who decide to explore performance assessments can benefit from the experience of those who have already worked in this area. A supplement lists 27 resources for those interested in performance assessment, as well as organizations active in the field and state contact people. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Indiana Education Policy Center

POLICY BULLETIN

Bloomington Office

April 1992—No. PB-B13

ED360388

Performance Assessment

by
Mark Buechler

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

M. McCarthy

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Buechler is a Research Associate at the Indiana Education Policy Center Bloomington Office.

To supplement or replace multiple-choice tests, educators in many states are devising assessments that require students to produce rather than reproduce knowledge.

To be promoted to Mount Olympus, Hercules had to complete 12 labors. If he were a student in a contemporary American classroom, however, he would probably have to take a multiple-choice test to stake his claim to immortality, filling in tiny circles on an answer sheet instead of obtaining golden apples or rescuing sacred beasts.

But in some states, Hercules would get to demonstrate his ability by performing tasks rather than answering multiple-choice questions. Such demonstrations, or performance assessments, are being touted not only as a better way to foster and measure student achievement but also as a lever for school restructuring.

Frustration with Multiple-Choice Tests

At the root of the performance assessment movement is a fairly widespread dissatisfaction with high-stakes multiple-choice tests. In this era of accountability, policymakers and the general public are clamoring for more and more information on student achievement. Because multiple-choice tests are easy to administer, relatively inexpensive, and seemingly objective, they are by far the most popular means of large-scale assessment.

However, such tests are beset with problems, say critics. "Rather than being 'objective' instruments, standardized tests often produce results that are *inaccurate, inconsistent and biased* against minority, female and low-income students," write Noe Medina and Monty Neill (p. 3). What's more, critics declare, the tests are not a credible gauge of students' ability to solve problems, read, or write, much less to think creatively or work together. Finally, critics claim that multiple-choice tests corrupt the teaching process. Pressured to produce high test scores, teachers narrow instruction toward the limited skills targeted by the tests.

Performance Assessment

It is unrealistic to expect teachers to change instructional methods in the face of tests that deter such changes. For instruction to improve, the tests themselves will have to change, say many multiple-choice test critics. Thus, educators in some states are developing a new generation of tests called performance assessments, or authentic assessments, in which students demonstrate mastery by executing tasks or creating products rather than by selecting responses from a fixed menu. Performance assessments may involve:

- *Uniform tasks*, such as timed essay questions or open-ended math problems, which students take statewide at a designated time.
- *Embedded tasks*, or tasks incorporated into the classroom setting, which call on students to frame problems, conduct research, design science experiments, write reports, give presentations, converse with teachers, or make videotapes.

TM020346

- *Portfolios*, or systematic collections of student work, including a wide range of finished products, early drafts, teachers' comments, and students' reflections on their own work. Portfolios enable teachers to evaluate student growth as well as discrete achievements.

Uniform tasks are often developed by states or testing companies with input from teachers, while embedded tasks and portfolios are often developed by teachers in collaboration with experts.

To be sure, many teachers have always used some version of performance assessment in the classroom. English teachers have students write, for example. One of the purposes of the performance assessment movement is to get more teachers to do the same. But advocates look beyond the individual classroom toward school, district, and state assessments based on performance. They hope to devise tests that will meet the public's demand for accountability even while improving classroom instruction.

SAMPLE PERFORMANCE ASSESSMENT

What Does It Cost to Take a Shower?

Part 1: Getting Started by Yourself

Think of five energy costs involved in taking a shower and identify the information needed to estimate these costs.

Part 2: Group Work

(a) How much energy and water is needed to take a shower? What are the costs of these components? Design and carry out a study to answer these questions as completely as possible.

(b) Write a report of your investigation. Make certain you address all of the objectives on the Objectives Rating Form.

(c) Present the results of your study to the class, including visual materials to make your presentation clearer.

Part 3: Finishing by Yourself

Using the data from Part 2, compute the savings in a single shower that would be achieved by (a) reducing the rate of water flow by 30% and (b) lowering the water temperature by 2°C.

A task (abridged) from Connecticut's Common Core of Learning Performance Assessment Project, sponsored by the National Science Foundation.

Schools and Districts

Hundreds of schools around the country are already experimenting with performance assessments. One of the most acclaimed is IPS Key School in Indianapolis, where videotapes of student projects form the principal means of evaluation. Schools belonging to the Coalition of Essential Schools, such as Walden III in Wisconsin and Central Park East in New York, require students to demonstrate mastery through an "Exhibition" before they graduate.

In some cases, whole districts have committed to performance assessment. In Pittsburgh, for example, secondary school teachers, working with researchers from the Educational Testing Service and Harvard University, have developed a districtwide portfolio process called ARTS PROPEL to document student learning in writing, music, and art. New standards, curricula, and teaching practices are being developed hand-in-hand with the portfolio process.

Statewide Performance Assessments

Many states are also experimenting with performance assessment. For example:

- All 4th graders in New York take a hands-on science test, rotating through five stations and using real science equipment to solve problems.
- California has added open-ended math problems to its 12th grade math test and a timed essay to its 8th and 12th grade language arts tests and is resolutely field testing new assessments in other subjects despite severe budget constraints.
- Indiana (and over two dozen other states) also uses a timed essay as part of its statewide assessment program.
- Arizona and Maryland conduct a *two-day* writing assessment that lets students prewrite and draft on the first day and rewrite on the second. Both states are also developing performance tests in reading, math, and other subjects.
- Kentucky is committed to implementing a multi-subject, multi-grade performance assessment system by 1995.

The two states farthest along the performance assessment trail are Connecticut and Vermont. Supported by a $1 million grant from the National Science Foundation, Connecticut teachers and researchers are spearheading a multi-state compact of schools that assess sustained performance tasks in science and math. These tasks may take students weeks or months to complete and may involve group as well as individual work. Teachers evaluate not only the products that result but also student attributes such as curiosity and persistence. Connecticut's long-range goal is to

The views expressed in this publication are those of the author and do not necessarily represent positions of the Indiana Education Policy Center or its funders, the Lilly Endowment, Inc., and Indiana University.

© 1992 *Indiana Education Policy Center*

BEST COPY AVAILABLE

incorporate performance assessment into its statewide testing program.

Vermont is the first state in the nation to use portfolios as part of a statewide assessment program. Students in 4th and 8th grade compile portfolios in writing and math. The Vermont Department of Education has established minimum requirements for portfolio content, but individual teachers have ample leeway as to what to include. Teachers grade their own students' portfolios based on state standards. To ensure consistency in scoring, random samples of portfolios are selected for grading by a second teacher. If significant discrepancies arise, state-trained assessors adjudicate. Vermont's program also has two other components: a "best piece" chosen by each student and a uniform task administered statewide.

Authentic assessments replicate the challenges that typically face businesspeople, scientists, or historians.
—Grant Wiggins

Benefits

Performance assessments are *real*, advocates say. "Authentic assessments replicate the challenges and standards of performance that typically face writers, businesspeople, scientists, community leaders, or historians," says Grant Wiggins (pp. 703-704). As such, they give direct evidence of higher-order thought processes rather than the circumstantial evidence provided by multiple-choice tests.

Just as important as what performance assessments tell educators about students is what they do for the students themselves: offer genuine intellectual challenges. "Participation in authentic tasks is more likely to motivate students," says Fred Newmann (p. 460). Particularly in portfolio programs, students also learn to reflect on what they produce and to revise accordingly—habits of thought crucial to success in the real world.

Finally, performance assessment may be a way to restructure the schools—in the reverse order of the current approach. As things stand, many reformers change the organizational structure of a school first and expect changes in curriculum and instruction to follow. However, "organizational changes alone will not modify long-standing conceptions of knowledge and deeply socialized habits of didactic teaching," says Newmann (p. 463). If we first decide what abilities students should acquire, then give them the chance to demonstrate those abilities, and finally develop modes of organization and instruction to facilitate the demonstrations, "restructuring could form a new chapter in educational history" (p. 463).

Problems

Despite such glowing tributes, performance assessment is not without its share of problems. Among them:

Scoring. Whatever bias may creep into the content of multiple-choice tests, the scoring is unyieldingly objective. Critics of performance assessment (including many parents) worry that evaluation of open-ended tasks will let subjectivity slip back into the grading process, opening a whole new Pandora's Box of unfairness and unreliability.

Assessment developers have evolved a three-point strategy to address scoring concerns:

- Formulate explicit, systematic criteria for tasks.
- Make sure those rating the tasks are thoroughly trained.
- Use multiple raters and monitor results to ensure that criteria are applied consistently.

Validity. Validity is a technical term referring to the extent to which a test actually measures what it is supposed to measure. Psychometricians use sophisticated techniques for increasing the validity of multiple-choice tests, but similar techniques have yet to be developed for performance assessments. "What should especially trouble policy makers," says Gregory Cizek, is that "we have begun a search for genuine-*looking*, authentic-*looking*, real-*looking* assessments and have eschewed more rigorous standards of validity" (p. 699).

Multiple-choice tests may be superior in terms of technical validity, but that does not compensate for other flaws, says Dale Carlson, director of the California Assessment Program: "The narrowing of the curriculum that is taking place, inadvertently, as a result of our not [using performance assessments] is in the long run more damaging than any statistical problems that may be occurring in the short run" (in Rothman, p. 10). Other proponents assert that the superiority of performance assessment in terms of "ecological validity"—fidelity to the overall goals of the curriculum—more than makes up for statistical deficiencies.

*We have begun a search for genuine-*looking *assessments and have eschewed more rigorous standards of validity.*
—Gregory Cizek

Instruction vs. Accountability. It often seems that performance assessments with the greatest potential for improving instruction—portfolios and extended tasks—are least amenable to aggregate scoring (which enables comparison across schools), while performance assessments that are easiest to aggregate—uniform, state administered tasks—are the least likely to improve instruction. "There is no question that if we wanted to assess students for the purposes of improving the quality of instruction, we could do it tomorrow," says Wiggins. "The problem is, for a variety of reasons, most people still want a technically rigid, comparative accountability system. Demanding that simultaneously is a real problem" (in Rothman, p. 10).

Time and Teacher Resistance. Performance assessments can impose a severe burden on teachers accustomed to preparing students for traditional tests. Old-fashioned drill-and-practice methods are inappropriate for authentic assessments, and many teachers have to revamp their classroom approach. Evaluating tasks also takes much more time than

simply grading answer sheets. "I thought, 'Oh God, here's another thing they're throwing at us,'" said one Vermont teacher (in Bandler and Holcombe, p. 15). Despite the new demands, though, teachers in Vermont and elsewhere, heartened by the enthusiastic responses of students, have for the most part reacted favorably to such programs.

Cost. Vermont has spent about $400,000 per year over three years for its writing and math program in grades 4 and 8; Kentucky is spending $3.5 million during the first year of development of its multi-subject, multi-grade assessments and plans to spend up to $30 million over the next five years. It costs less than a penny for a computer to grade an answer sheet, versus 80¢ to $5.00 for people to grade a uniform task. In short, performance assessments are expensive to develop and score, although the costs vary widely depending on the type and extent of assessment. Whatever the cost, experts say that a big portion of it must be budgeted for teacher training and planning time, especially in programs that use sustained tasks or portfolios.

Although expensive, performance assessment offers hidden benefits, proponents claim. It motivates students. It helps teachers refine standards for student achievement. And it provides some of the best staff development that teachers will ever get.

Suggestions

If a state decides to explore performance assessment, policymakers can profit from advice given repeatedly by those who have worked in the field:

- To the extent possible, involve teachers at every stage in the assessment process: development, scoring, revising.
- Allow for necessary increases in staff development and planning time.
- Start small, with one or two subjects in one or two grades, before moving to more extensive programs.

Perhaps most importantly, says Joan Baron, director of Connecticut's program, "Be patient." Giving developers time to wrestle with still-existing problems may ultimately determine whether performance assessment is just another educational fad or a lever for genuine school reform.

References

Bandler, J., & Holcombe, R. (1991, May 5). The portfolio problem. *Vermont Sunday Magazine*, pp. 5+.

Cizek, G. J. (1991, May). Innovation or enervation. *Phi Delta Kappan*, pp. 695-699.

Medina, N., & Neill, D. M. (1988). *Fallout from the testing explosion*. Cambridge, MA: FairTest.

Newmann, F. M. (1991, February). Linking restructuring to authentic student achievement. *Phi Delta Kappan*, pp. 458-463.

Rothman, R. (1990, September 12). New tests based on performance raise questions. *Education Week*, pp. 1+.

Wiggins, G. (1989, May). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, pp. 703-713.

POLICY BULLETIN NO. PB—B13
April 1992

Indiana Education Policy Center
Bloomington Office
Smith Center for Research
in Education, Suite 170
Indiana University
Bloomington, IN 47405
(812) 855-1240

Barry Bull, Co-director
Gayle Hall, Associate Director

The Indiana Education Policy Center is funded by the Lilly Endowment, Inc., and Indiana University to provide nonpartisan research on education issues to Indiana policymakers and other education stakeholders to improve public education.

NON-PROFIT ORG.
U.S. POSTAGE
PAID
Bloomington, IN 47405
Permit No. 2

Performance Assessment
Selected Resources

Articles, Reports, and Books

Alternative assessment. (1991). *Policy Briefs* (Number 15 & 16). (A collection of articles and updates on assessment activity in the north central region, published by the North Central Regional Educational Laboratory, 1900 Spring Road, Suite 300, Oak Brook, IL, 60521.)

Archbald, D. A., & Newmann, F. M. (1988). *Beyond standardized testing: Assessing authentic academic achievement in the secondary school.* Reston, VA: National Association of Secondary School Principals.

Arter, J. A. (1990). *Using portfolios in instruction and assessment.* Portland: Northwest Regional Educational Laboratory.

Aschbacher, P. R. (1991). *Alternative assessment: State activity, interest, and concerns* (CSE Technical Report 322). Los Angeles: Center for Research on Evaluation, Standards, and Student Testing.

Baron, J. B., Forgione, P. D., Rindone, D. A., Kruglanski, H., & Davey, B. (1989, March). *Toward a new generation of student outcome measures: Connecticut's common core of learning assessment.* Paper presented at the American Educational Research Association Annual Meeting, San Francisco.

Cizek, G. J. (1991, May). Innovation or enervation. *Phi Delta Kappan,* pp. 695-699.

Cizek, G. J. (1991, October). Confusion effusion: A rejoinder to Wiggins. *Phi Delta Kappan,* pp. 150-153.

Fiske, E. B., with Reed, S., & Sautter, R. C. (1991). *Smart schools, smart kids.* New York: Simon & Schuster. (See especially Chapter 5, "Beyond Testing.")

Gilbert, J. C. (1990). *Performance-based assessment resource guide.* Denver: Colorado Department of Education.

Haney, W., & Madaus, G. (1989, May). Searching for alternatives to standardized tests: Whys, whats, and whithers. *Phi Delta Kappan,* pp. 683-687.

Harrington-Lueker, D. (1991, April). Beyond multiple choice: The push to assess performance. *The Executive Educator,* pp. 20-22.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). *Complex, performance-based assessment: Expectations and validation criteria* (CSE Technical Report 331). Los Angeles: Center for Research on Evaluation, Standards, and Student Testing.

Maeroff, G. I. (1991, December). Assessing alternative assessment. *Phi Delta Kappan,* pp. 272-281.

Medina, N., & Neill, D. M. (1988). *Fallout from the testing explosion.* Cambridge, MA: FairTest.

Mills, Richard. (1989, December). Portfolios capture rich array of student performance. *The School Administrator,* pp. 8-11.

Mitchell, R. (1992). *Testing for learning: How new approaches to evaluation can improve American schools.* New York: Macmillan.

National Commission on Testing and Public Policy. (1990). *From gatekeeper to gateway: Transforming testing in America.* Chestnut Hill, MA: Author.

Newmann, F. M. (1991, February). Linking restructuring to authentic student achievement. *Phi Delta Kappan,* pp. 458-463.

Performances and exhibitions: The demonstration of mastery. (1990, March). *Horace, 6*(3). (Bulletin of the Coalition of Essential Schools, Box 1938, Brown University, Providence Rhode Island, 02912.)

Redirecting assessment [Special issue]. (1989, April). *Educational Leadership, 46*(7). (The entire issue is devoted to alternative assessment. It includes articles by Joan Baron, George Madaus, Ed Roeber, Grant Wiggins, Dennie Palmer Wolf, and others.)

Rothman, R. (1990, September 12). New tests based on performance raise questions. *Education Week,* pp. 1+.

Valencia, S. (1990, January). A portfolio approach to classroom reading assessment: The whys, whats, and hows. *The Reading Teacher,* pp. 338-340.

Vermont Department of Education. (1991). *Looking beyond "the answer": The report of Vermont's mathematics portfolio assessment program.* Montpelier: Author.

Vermont Department of Education. (1991). *"This is my best": The report of Vermont's writing assessment program.* Montpelier: Author.

Wiggins, G. (1989, May). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan,* pp. 703-713.

Wiggins, G. (1990, December). *The case for authentic assessment* (EDO-TM-90-10). Washington, DC: ERIC Clearinghouse on Tests, Measurement, and Evaluation.

Wiggins, G. (1991, May). A response to Cizek. *Phi Delta Kappan,* pp. 700-703.

Organizations Involved in Performance Assessment Research or Development

Center for Research on Evaluation, Standards, and Student Testing (CRESST)
Eva Baker and Robert Linn, Co-directors
Contact person: Ron Dietel, Director of Communications
UCLA Graduate School of Education
405 Hilgard Avenue
145 Moore Hall
Los Angeles, CA 90024
Phone: 310-206-1532

Center on Learning, Assessment, and School Structure (CLASS)
Grant Wiggins and Holly Houston, Directors
39 Main Street
Geneseo, NY 14454
Phone: 716-243-5500

Council for Basic Education
Ruth Mitchell, Associate Director
725 Fifteenth Street, N.W.
Washington, DC 20077
Phone: 202-347-4171

Educational Testing Service (ETS)
Mary Fowles, Principal Measurement Specialist
P-O6
Rosedale Road
Princeton, NJ 08541
Phone: 609-734-5228

National Commission on Testing and Public Policy
George Madaus, Executive Director
McGuinn Hall, Room 531
Boston College
Chestnut Hill, MA 02167
Phone: 617-552-4521

Northwest Evaluation Association
Allan Olson, Executive Director
5 Centerpointe Dr., Suite 100
Lake Oswego, OR 97035
Phone: 503-624-1951

PACE (Performance Assessment Collaboratives for Education)
Dennis Palmer Wolf, Director
8 Story Street
Cambridge, MA 02138
Phone: 617-496-2770

Portfolio Assessment Clearinghouse
Winfield Cooper and Jon Davies, Co-directors
c/o San Dieguito Union High School District
710 Encinitas Blvd.
Encinitas, CA 92024
Phone: 619-753-6491

Student Assessment Consortium
Council of Chief State School Officers (in collaboration with CRESST)
Ed Roeber, Director
One Massachusetts Avenue, N.W., Suite 700
Washington, DC 20001-1431
Phone: 202-408-5505

State Contact People

Joan Baron
Common Core of Learning Assessment Coordinator
Connecticut State Department of Education
165 Capital Avenue, Room 340
Hartford, CT 06106
Phone: 203-566-5454
(*Note*: A comprehensive monograph on Connecticut's progress in developing performance assessments is scheduled for publication in the summer of 1992.)

Dale Carlson
Assistant Superintendent of Public Instruction
California Assessment Program
California Department of Education
PO Box 944272
721 Capitol Mall
Sacramento, CA 95816
Phone: 916-657-3011

Geof Hewitt, Writing Consultant
Robert Kenney, Mathematics Consultant
Vermont Portfolio Assessment Program
Vermont Department of Education
120 State Street
Montpelier, VT 05620
Phone: 802-828-3111

Indiana Education Policy Center
Bloomington Office
School of Education
Suite 326
Indiana University
Bloomington, IN 47405